临床医生如何利用回顾性临床资料进行回顾性队列研究

王小钦

复旦大学循证医学中心

复旦大学附属华山医院血液科

电话：021-52888252，Email：wangxiaoqin@shmu.edu.cn

【摘要】回顾性临床病例资源丰富，但是存在的一些问题限制了临床医师利用这些数据开展临床研究，例如基线资料不全、非随机分组导致混杂因素多、失访率高等问题。采用限制、配对、分层、多因素分析、倾向性评分等多种方法可以克服回顾性资料的固有缺陷，得出可靠的结论。

【关键词】临床研究；回顾性队列研究；混杂因素；倾向性评分

How to use retrospective clinical data to do retrospective cohort study?

Wang Xiao-qin

Evidence-based Medicine Centre of Fudan University, Hematology department of Huashan Hospital, Fudan University

Telephone：021-52888252，Email: wangxiaoqin@shmu.edu.cn

[Abstract] Retrospective clinical data are abundant, but there are some problems that limit clinicians' use of these data for clinical research, such as incomplete baseline data, confounding bias due to nonrandom grouping, and higher loss of follow-up. The inherent defects of retrospective design can be overcome by the limitations, matching, stratification, multivariate analysis, propensity score and other methods.

[Key words] Clinical Research; Retrospective Cohort study; Confounding factors; Propensity score analysis

回顾性临床病例资料数据庞大，资源丰富，如果临床医师可以合理利用它们进行临床研究，则可以节约大量的时间和经费，很快得出结论和成果。但是回顾性数据存在基线数据不全、不详细，治疗方案不统一，随访间期不等，失访率高等问题，所以需要良好的研究设计和统计分析来克服这些缺陷，才能得出可靠的结论。

临床研究根据有无人为设计的干预因素分为实验性研究和观察性研究，观察性研究又根据有无对照组，分为描述性和分析性研究，队列研究、病例-对照研究和横断面研究是常用的分析性研究设计方法[1]。除了不能进行实验性研究和前瞻性队列研究以外，回顾性资料可以进行病例报道和病例分析，也可以进行病例-对照研究、回顾性队列研究和横断面研究，所以可以设计的研究类型是非常多的。例如罕见病可以进行病例报道（<10 例 ，≥10 例的病例可以进行病例分析。病因学研究多采用病例-对照的方法进行设计，疗效评价、预后研究和病因学研究可以采用回顾性队列研究的设计，诊断试验评价可以采用横断面研究的方法。由于回顾性队列研究应用范围最广，所以下面重点介绍如何进行回顾性队列研究的设计和统计分析，以期达到结论真实可靠的目的。

1 回顾性队列研究设计阶段需要注意的几个问题

1.1 提出 PICO 问题

回顾性队列研究的基本步骤包括：明确研究目的，确立结局指标和研究因素，提出 PICO 问题（研究对象 P、干预或暴露因素 I、对照组 C、结局指标 O）；确定研究对象的纳入和排除标准；计算样本量；收集病例的基线资料；随访病例结局；进行统计分析，校正各种混杂因素；总结、撰写论文。

写出 PICO 问题可以帮助研究者理清思路和明确研究目的，是非常重要的。例如研究 A 基因突变是否是急性髓系白血病复发的高危因素，P（研究对象）是某种类型的急性髓系白血病，I（暴露因素）是有 A 基因突变，C（对照组）是无 A 基因突变，O（结局指标）是白血病复发，在统计分析阶段需要校正的因素有年龄、白细胞计数、原始细胞计数、染色体核型、化疗方案等影响结局的因素。

1.2 基线数据缺失问题

针对基线资料不全这个缺陷，在设计课题阶段可以采用限制的方法，即设立纳入和排除标准，把关键资料缺失的病例排除。为了样本有代表性，一般设立一个病例纳入的年限，如纳入 2015 年 1 月 1 日～2018 年 12 月 31 日的符合纳入条件的所有病例（连续样本），而且排除标准不能太严格。如果排除的病例较多，最好进行一个缺失数据模式分析，如果是完全随机缺失或随机缺失，删去缺失病例对结果的影响较小。也可以比较纳入研究的对象和剔除的对象的基线临床特征，如果无临床特征的显著差异，可以说明纳入病例的代表性较好，基本可以代表总

体病例。

1.3 样本量的计算

回顾性分析是否需要计算样本量呢？只要有对照组，需要进行统计学分析、计算 P 值的研究都需要计算样本量，因为样本量太小，有可能得到假阴性的结果。样本量的计算取决于主要结局指标[2]，例如研究某一疾病的缓解率、无进展生存率、总生存率的样本量是完全不同的，样本量依次递增。所以研究中设计主要结局指标和次要结局指标很重要。

1.4 配对的方法

队列研究中，可以把没有某暴露因素（或治疗方法）的所有病例作为对照组，也可以在设计阶段，采用配对的方法进行对照组的选择，例如进行年龄、重要的预后因素（疾病分期、亚型、并发症等）配对，根据不同的疾病而定，使两组可比性较好。如果应用手工配对的方法，一般配对的因素为 2-4 个，否则很难配到合适的对照。如果应用一些统计方法进行配对，如下面案例应用的迭代扩展半径法（iterative expanding radius matching），可以适当增加配对的因素。配对的比例一般是 1:1～1:4 配对，超过 4 以后，再增加配对例数，统计效能(Statistical Power)增加不多。

案例分析：发表在新英格兰医学杂志的一项多中心、回顾性队列研究“应用 HLA 不相合活体供体进行肾移植有生存获益”[3]，其目的是研究与单纯等待的患者（或者等待后进行死者供体肾移植患者）相比，HLA 不相合的活体供体肾移植是否有生存获益。数据来源于美国移植登记数据库，其中有 1025 例 HLA 不相合的活体肾移植病例，对照组采用 1:5 配对，配对因素为年龄、性别、种族、既往进行肾移植的次数、HLA 抗体水平、有无糖尿病、肾脏替代治疗时间 7 个因素。配对方法采用迭代扩展半径法。从单纯等待未进行肾移植的患者中选择 5125 例作为对照组 1，从等待后进行了 HLA 相合死亡供体肾移植中选择 5125 例作为对照组 2（表 1）。然后进行统计分析，分析总体生存时间的差异，发现与等待患者相比，及早应用 HLA 不相合的活体肾供体可以提高生存时间。

表 1．HLA 不相合活体肾移植患者与配对的对照组的临床特征比较

	HLA 不相合活体肾移植患者	单纯等待未进行肾移植的对照组（N=5125）	等待后进行了死亡供体肾移植的对照组

	(N=1025)		(N=5125)
移植时年龄（岁）	45.0±12.8	45.9±12.1	46.6±12.0
女性患者-例数（%）	687（67.0）	3444（67.2）	3452（67.4）
黑人患者-例数（%）	168（16.4）	844（16.5）	840（16.4）
既往进行过肾移植的次数			
-例数（%）			
0	612（59.7）	3323（64.8）	3321（64.8）
1	350（34.1）	1584（30.9）	1552（30.4）
2	57（5.6）	202（3.9）	231（4.5）
≥3	6（0.6）	16（0.3）	21（0.4）
中位抗体水平(%)-中位数	66（22-94）	68（20-95）	68（21-95）
（四分位数）			
糖尿病-例数（%）	210（20.5）	1047（20.4）	1047（20.4）
肾脏替代治疗时间（年）	7.3±7.6	7.3±7.4	7.4±7.4

2. 回顾性队列研究在分析阶段需要注意的几个问题

在统计分析阶段可以采用分层分析、多因素分析、倾向性评分等多种方法校正各种混杂因素，以明确研究因素是否独立危险因素或预后因素。

2.1 分层分析

在分析阶段，对一些最重要的影响结局、预后的因素进行分层分析，目的是观察消除该影响因素以后，两组结局比较是否有变化。例如上文中，对 HLA 抗体阳性的不同水平进行分层分析，发现不同的抗体水平下，HLA 不相合活体供体肾移植仍有生存获益。

案例分析：在一项比较输血依赖的骨髓增生异常综合征应用去铁治疗和不去铁治疗是否生存时间不同的研究中[4]，结论是去铁治疗可以延长生存期。因为输血量是疾病严重程度的一个标志，疾病越严重每个月输血量越多，生存时间越短。所以对每个月输血的量（≤3 个单位红细胞和>3 个单位红细胞）进行分层分析，发现不同输血量情况下，去铁治疗均可以延长生存期（图 2），更证明了结论的可靠性。

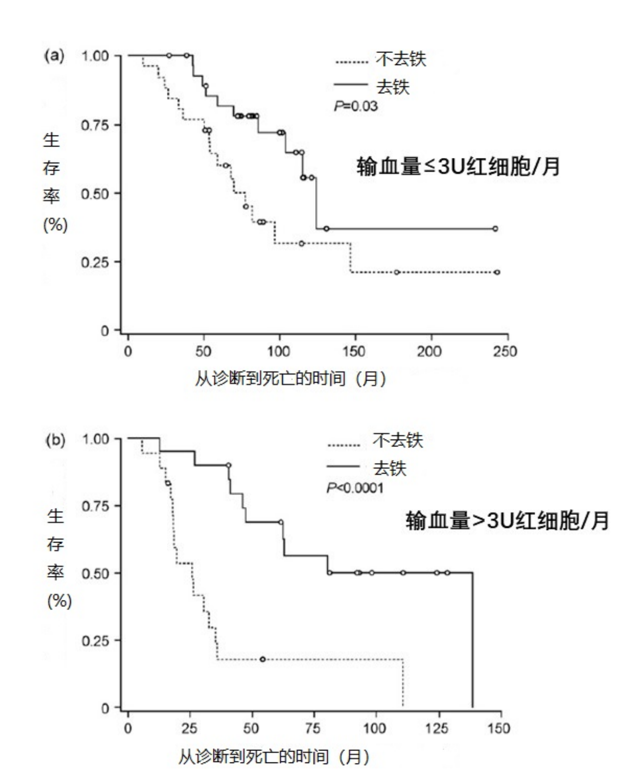


图 2．去铁治疗和不去铁治疗对骨髓增生异常综合征生存时间的影响

注：不同的输血量作为分层因素，(a)输血量≤3 个单位红细胞/月，(b)输血量>3 个单位红细胞/月

2.2 多因素分析

因为影响结局的混杂因素较多，多因素分析可以同时校正多个混杂因素，所以多因素分析是队列研究中常用的统计方法。常用的多因素分析方法有多元回归分析、Logistic 回归分析、Cox 回归模型。

这三种方法如何选择呢？当结局变量为连续变量时，选择多元回归分析。当结局变量为分类变量，同时又要考虑时间对该结局变量的影响（有时间变量），可以选择 Cox 回归模型。当结局变量为分类变量，不考虑时间变量时，可以选择 Logistic 回归分析。

案例分析：在一项“老年贫血患者应用促红细胞生成素的治疗反应和心血管结局”回顾性队列研究中[5]，应用 Logistic 回归分析应用促红细胞生成素后治疗反应的影响因素，应用 COX 回归分析应用促红细胞生成素后心血管事件的发生结局，研究应用促红细胞生成素后是否增加心血管事件的发生。治疗反应的评价是统一在治疗 2 个月后进行评价，所以不需要考虑时间变量。而应用促红细胞生

成素后心血管事件的发生需要考虑时间变量，有的患者使用后很短时间就发生了心血管事件，有的患者很久以后才发生心血管事件。研究结论是：不论是不进行校正的单因素分析[HR=1.21(95%CI：0.88-1.66)，P=0.234]，还是用5个因素(存在其他血细胞减少、血红蛋白水平、年龄、性别、并发症指数）进行校正的多因素分析[HR=0.94(95%CI：0.67-1.31)，P=0.693]，均认为应用促红细胞生成素后并不增加心血管事件的发生危险。该文根据不同的结局变量，选择了2种多因素分析的方法。

如何选择进入多因素分析的变量呢？从文献或临床经验中确定需要进行校正的因素，也就是可能会影响结局的变量，可以进行单因素分析进行初步筛选，或者不进行单因素分析，全部纳入多因素分析。选择变量时主要考虑临床意义，但是如果样本量有限，要控制进入多因素分析的变量数量，先进行单因素筛选也是可行的。上述骨髓增生异常综合征应用去铁治疗提高生存时间的研究中[4]，采用先进行单因素分析，把单因素分析有统计学意义的变量纳入多因素 COX 回归分析这种方法。而老年贫血应用促红细胞生成素治疗的文章中[5]，选择的5个校正变量，是根据临床意义选择的，未经过单因素分析筛选。在我们自己的研究中，根据实际情况，可以选择其中一种方法，或者 2 种方法联合使用，同时考虑单因素分析的结果，也纳入单因素分析没有意义，但是很有临床意义的变量。

2.3 倾向性评分法

倾向性评分方法可以在分析阶段有效平衡非随机对照研究中的混杂偏倚，使研究结果接近随机对照研究(RCT)的效果，所以有人称其为事后随机化方法。可以应用倾向性评分进行配对、分层、回归等分析方法来校正混杂因素，最常用的是倾向性评分配对方法。其与普通配对方法的不同是倾向性评分方法可以同时进行多个变量的配对，而普通配对方法的配对变量是有限的。

案例分析：费城染色体阳性的急性淋巴细胞白血病（ALL）应用 Hyper-CVAD 化疗方案+帕纳替尼与 Hyper-CVAD+达沙替尼治疗相比较，哪一个治疗反应、无事件生存、总体生存率更好呢？因为两组基线临床特征有一定的差异，作者采用倾向性评分配对的方法进行了研究[6]。首先选择配对的8个因素：年龄、体能状态、白细胞计数、染色体危险度分组、BCR-ABL1 的转录类型、是否应用利妥昔单抗、CD20阳性的原始细胞比例、是否有中枢神经系统累及。然后应用 Logistics

回归方法计算这8个因素的倾向分值。再应用最近邻居方法对评分进行1:1配对，原有110例ALL病例，经过倾向性评分配对后，有41对（82例）病例配对成功。最后，对这82例病例进行统计分析，发现应用帕纳替尼组无事件生存和总体生存率均优于达沙替尼组。

倾向性评分方法能否很好控制混杂因素取决于纳入哪些协变量计算倾向分值，目前尚无公认的标准，但是一般认为应该纳入所有的、已知的、可能影响结局的所有混杂因素。

2.4 失访率问题

回顾性队列研究有失访是不可避免的，一般要求失访率控制在10%之内，最多不能超过20%，否则会存在选择性偏倚，影响结果。所以，要尽量控制失访。如果有较多的病例失访，可以进行敏感性分析，把失访病例的结局分别按照最差结局和最好结局计算，如果结论没有显著差异，认为失访对结论的影响较小，结论比较可靠。否则，结论的可靠性较差。

3. 撰写论文规范

队列研究属于观察性研究，撰写论文时要按照观察性研究报告规范（Strengthening the Reporting of Observational Studies in Epidemiology，STROBE)来进行论文的写作。详细的STROBE声明清单见www.strobe-statement.org网站。在进行研究设计之前应该先阅读该声明，按照清单上的内容进行设计，才能保证研究质量和论文质量。

4. 总结

临床医生可以利用回顾性资料进行各种研究设计，根据研究目的的不同选择不同的研究方法，回顾性队列研究是常用的设计方法。两组不均衡可比，存在很多的混杂因素是回顾性队列研究中比较突出的问题，可以应用上述介绍的各种方法来校正混杂因素，同一研究中可以同时应用多种方法来进行校正。

参考文献：

1. Schulz KF, Grimes DA. 王吉耀，译. 《柳叶刀》临床研究基本概念[M]. 北京：人民卫生出版社. 2010:4-9

2. 曾于珍，陈世耀. 临床研究结局指标选择与样本量估计[J]. 协和医学杂志，

2018, 9（1）:87-92

3. Orandi BJ, Luo X, Massie AB, et al. Survival Benefit with Kidney Transplants from HLA-Incompatible Live Donors[J]. N Engl J Med. 2016; 374(10): 940–950. doi:10.1056/NEJMoa1508380.

4. Rose C, Brechignac S, Vassilief D, et al. Does iron chelation therapy improve survival in regularly transfused lower risk MDS patients? A multicenter study by the GFM (Groupe Francophone des Myélodysplasies) [J]. Leuk Res. 2010;34(7):864-70. doi: 10.1016/j.leukres.2009.12.004.

5. Gowanlock Z, Sriram S, Martin A, et al. Erythropoiesis-stimulating agents in elderly patients with anemia: response and cardiovascular outcomes [J]. Blood Adv. 2017;1(19):1538-1545. doi: 10.1182/bloodadvances.2017007559.

6. Sasaki K, Jabbour EJ, Ravandi F, et al. Hyper-CVAD plus ponatinib versus hyper-CVAD plus dasatinib as frontline therapy for patients with Philadelphia chromosome-positive acute lymphoblastic leukemia: A propensity score analysis[J]. Cancer. 2016;122(23):3650-3656. doi: 10.1002/cncr.30231.